DOCUMENT RESUME

ED 202 879 TM 810 300

AUTHOR Abrami, Philip C.; Murphy, Vincent
TITLE A Catalogue of Systems for Student Ratings of Instruction.
INSTITUTION McGill Univ., Montreal (Quebec). Centre for Teaching and Learning Services.
PUB DATE Jun 80
NOTE 63p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Attitude Measures; *Course Evaluation; Educational Quality; Higher Education; *Rating Scales; *Student Attitudes; *Student Evaluation of Teacher Performance

ABSTRACT

This catalogue briefly describes the following 12 systems for student ratings of instruction in higher education: (1) Purdue Cafeteria System (Cafeteria); (2) Course Faculty Instrument (CFI); (3) Arizona Course/Instructor Evaluation Questionnaire (CIEQ); (4) Endeavor Instructional Rating System (Endeavor); (5) University of Washington Instructional Assessment System (IAS); (6) Instructor and Course Evaluation (ICE); (7) Instructor and Course Evaluation System (ICES); (8) Instructional Development and Effectiveness Assessment System (IDEA); (9) Instructor Designed Questionnaire (IDQ); (10) Student Instructional Report (SIR); (11) Student Instructional Rating System (SIRS); and (12) Student Perception of Teaching (SPOT). The information provided is intended to be used as a starting point for faculty who are primarly interested in adopting one of these systems, but it may also be helpful as an aid to those designing local systems to meet unique needs. Selection criteria includes attention to careful development and empirical validation. Information typically included for each system consists of the title, contact, format, results, special features, development and validation, and references. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

CTLS
Centre for Teaching and Learning Services
McGill University

ED 232879

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A CATALOGUE OF SYSTEMS FOR STUDENT RATINGS OF INSTRUCTION

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
P.C. Abrami
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Philip C. Abrami

and

Vincent Murphy

TM 810 300

Campus address:
Room 407, Macdonald-Harrington
(Old Chemistry) Building,
Telephone (514) 392-8320

Postal address:
815 Sherbrooke Street West,
Montreal, Quebec
Canada H3A 2K6

A CATALOGUE OF SYSTEMS FOR STUDENT RATINGS OF INSTRUCTION

Philip C. Abrami
and
Vincent Murphy

Centre for Teaching and Learning Services
McGill University
815 Sherbrooke Street West
Montreal, Quebec
Canada H3A 2K6

© June 1980

Table of Contents

Introduction	iii
The Purdue Cafeteria System (Cafeteria)	1
Course Faculty Instrument (CFI)	5
Arizona Course/Instructor Evaluation Questionnaire (CIEQ)	9
The Endeavor Instructional Rating System (Endeavor)	14
The University of Washington Instructional Assessment System (IAS)	18
Instructor and Course Evalaution (ICE)	23
Instructor and Course Evaluation System (ICES)	28
Instructional Development and Effectiveness Assessment System (IDEA)	33
Instructor Designed Questionnaire (IDQ)	40
The Student Instructional Report (SIR)	43
Student Instructional Rating System (SIRS)	48
Student Perception of Teaching (SPOT)	52

Introduction

This catalogue briefly describes twelve systems for student ratings of instruction in higher education. The information provided is intended to be used as a starting point for faculty who are primarily interested in adopting one of these systems but it may also be helpful as an aid to those designing local systems to meet unique needs.

Attention to careful development and empirical validation were criteria used in selecting systems for inclusion in the catalogue. We also attempted to include systems which are in fairly wide use and thus represent fairly common, accepted approaches to teaching and learning and its evaluation. In addition, some systems are readily adaptable to unusual teaching situations, for example, studio courses, computer-based instruction, etc. Finally, we use the term "system" because this catalogue represents more than a description of teacher rating forms. The systems often include detailed procedures for the analysis and interpretation of rating data as well.

We rejected literally dozens of locally developed and/or very specific student rating systems (usually just rating forms). We did so not always because of concerns about their quality but, in many cases, because their focus seemed too narrow. Yet these narrowly focused systems may be most appropriate in meeting local needs. Lists of these systems, including brief descriptive summaries and sample rating forms, are available from the Centre for Teaching and Learning Services.

As noted, we have attempted to include rating systems which meet certain minimum standards for quality and utility.

Nevertheless, the listing of a system in the catalogue does not constitute our unqualified endorsement of it. We do not claim that any system meets all standards--psychometric, cost effectiveness, or whatever--which others might apply to it. In part, the setting of such standards is a matter of personal judgment dependent on local conditions and needs. Consequently, we urge potential users to familiarize themselves with the relevant original reports and descriptions of a system. In addition, we recommend that users acquaint themselves with the literature on instructional evaluation in general.

The task of collecting the many available rating systems was aided by the cooperation of our colleagues in instructional evaluation and development. In particular, we would like to thank the members of two AERA special interest groups, Instructional Evaluation and Faculty and Instructional Development, and the directors of faculty development centers around North America who responded to our mail surveys.

To facilitate the use of the catalogue, we have limited our description of each system to a few pages and described the major features of each system under seven headings. While we believe there are similarities between the systems, there are also substantial differences among them in development, application, flexibility, cost, support services, etc. We believe these differences offer the potential user a true choice in adopting a student rating system.

The following information is typically included under each heading:

Title

A complete title for the system and an acronym are given.

Contact

Name, address, and telephone number of the individual(s) responsible for user inquiries about the rating system are provided.

Format

This section provides a general description of the form(s) available with the system. This description typically includes information on: type, number, and organization of evaluation items, response alternatives, answer sheet format, student background questions, optional items, and open-ended questions.

Results

The Results summarize the data analyses provided by the system. Where possible, the summary describes the organization and general format of the output, statistics provided (e.g., means, standard deviations, factor scores, frequency distributions, histograms), availability and type of comparative data, etc.

Special Features

Under this heading are included any unique or outstanding features of the system not described elsewhere. Examples of special features include: the availability of interpretive manuals, newsletters, item bank catalogues, summary reports, and special consultants; the cost of forms and processing; and the existence of forms in foreign languages.

Development & Validation

This section provides brief synopses of published and unpublished research reports on the rating instruments used in each system. The studies surveyed may describe: procedures for selecting and organizing rating items; reliability estimates; test-criterion relationships; and the impact of student, teacher, course, and administration factors on rating form results.

Results

A list of the relevant articles particular to each system and emphasizing research evidence are given.

One final note. The Centre for Teaching and Learning Services maintains a complete file on each rating system for local users who desire additional information (e.g., sample rating forms, research reports). We urge other users to address inquiries to the appropriate system contact person.

Title: The Purdue Cafeteria System (Cafeteria)

Contact: Jerry O. Derry
Measurement and Research Center
Purdue Universty
West Lafayette, Indiana 47906
(317) 749-2207

Format

The Cafeteria system allows instructors to select up to 35 items from an item catalogue. It is also possible for instructors to include up to three items of their own design among the 35. These 35 items supplement a core set of five rating items which assess teacher-produced student motivation, instructor clarity, quality of assignments, overall course quality, and overall instructor quality. It is also possible for a department to make certain catalogue items mandatory for its faculty as part of the 35 noncore items. There is also a core set of five student background questions: class level, sex, expected grade, major, and reason for taking the course. All items except student background questions are scored on a five point scale ranging from strongly agree to strongly disagree. Individualized rating forms are printed by computer and allow space for student written comments. Cafeteria answer sheets are machine scorable.

Results

The results are typically returned directly to the instructor and are reported on a computer output which lists the items, the number of students choosing each item response

alternative, item medians, and percentile ranks for all catalogue-selected and core items. Frequency distributions for each student background question are also shown. Summary reports which provide normative data for each catalogue item are also available. These list the number selecting the item, a frequency distribution of item medians, percentile equivalents, and a frequency histogram.

Special Features

Cafeteria was designed for installation at other institutions with computers having FORTRAN capabilties. The one time system cost is $800 in U.S. funds and includes four FORTRAN program a 200 page operations manual, a computer managed catalogue of 200 items, and a norm library. The Cafeteria item catalogue is divided into 20 sections including: Clarity and Effectiveness of Presentation, Student Interest/Involvement in Learning, Broadening Student Outlook, etc. As an option, institutions which adopt the Cafeteria system may elect to augment the 200 Cafeteria items with an additional 100 locally developed items. Institutions may also choose to contract evaluations services with Purdue's Measurement and Research Center at a cost of approximately $.15 per rating form.

Development & Validation

The Cafeteria system represents the original attempt to provide a flexible, efficient, computer-based student evaluation system. The item catalogue was developed by first gathering a large number of items from other rating forms and editing them to conform to a single response format. These items plus others

written locally were then organized into content areas. After three semesters of pilot testing, the present Cafeteria system was designed. Derry (1977) reported that the average reliability of a Cafeteria rating form was .88. Derry, Seibert, Starry, Van Horn, and Wright (1974) suggested that the five core items appear to represent a unidimensional scale. Interitem correlations of item medians for 727 instructors ranged from .64 to .83.

References

Derry, J. O. Strengths and vulnerabilities of the Cafeteria model. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977. (Instructional Research Bulletin 77-1. West Lafayette, Indiana: Purdue University, Measurement and Research Center, 1977.)

Derry, J. O., Seibert, W. F., Starry, A. R., Van Horn, J. W., & Wright, G. L. Purdue instructor and course appraisal: the Cafeteria system: users manual. Lafayette, Indiana: Purdue Research Foundation, 1974.

Derry, J. O., Seibert, W. F., Starry, A. R., Van Horn, J. W., & Wright, G. L. The Cafeteria system: A new approach to course and instructor evaluation. Instructional Research Bulletin 74-1. West Lafayette, Indiana: Purdue University, Measurement and Research Center, 1974.

Title: Course Faculty Instrument (CFI)

Contact: Richard D. Freedman & Stephen A. Stumpf
New York University
Management/Organizational
Behavior Department
Graduate School of
Business Administration
100 Trinity Place
New York, New York 10006
(212) 285-6000

Format

The CFI rating form is a two-sided answer sheet which cannot be read directly by a computer scoring device. Part of the first page consists of questions dealing with student background information including major, program, reason for taking the course, work experience, GPA, and expected grade. A second part consists of seven items dealing with overall ratings of the course, text, and the instructor which are answered on a five-point scale. The second page consists of 51 questions listed under five headings: Instructor in Class, Instructor in General, Graded Assignments and Examinations, Texts and/or Required Reading, and Subject Matter. Response space is allotted for up to nine instructor designed questions at the end of this page. Responses to all items on the second page, including the instructor-designed questions are on a three point scale of Yes, ?, and No.

Results

The CFI measures seven dimensions of instruction: Instructor in Class, Instructor in General, Graded Assignments/Examinations, Subject Affect (students' interest in subject matter), Subject Difficulty, Subject Functionality (perceived usefulness of subject matter), and Texts and Other Reading. In addition, three overall evaluations are reported for decision-making purposes: an instructor rating, a course rating, and a text rating. How these results are tabulated and presented to the instructor is not stated.

Special Features

The CFI features open-ended response forms which provide diagnostic information to an instructor. These responses are not used in tabulating the overall rating of an instructor or a course and are seen only by the professor.

Development & Validation

Development and initial validation of the CFI was begun in 1976 (Freedman & Stumpf, 1978). Items were generated by asking 100 graduate business students for critical incidents, short descriptive phrases, or adjectives about effective or ineffective teaching performance on the five dimensions used as headings on the rating form. Each dimension was factor analyzed to exclude items within a dimension which did not contribute to the psychometric properties of the CFI. This resulted in the seven factorially simple dimensions for which results are provided. All items within each dimension correlated .39 or higher with a perceived learning criterion. All dimensions appear to be

reliable with a median coefficient alpha of .82, with a range of .72 to .85. The seven dimensions are not independent but intercorrelations are low, ranging from .07 to .55. The three-point response scale for CFI items was empirically compared to five-and-seven-point scales and found to be preferable. Stumpf, Freedman and Krieger (1979) demonstrated that the CFI is extendable to a variety of institutions by analyzing the CFI across five diverse populations. Factors such as the instructor's sex, rank, class size, etc. were found to be unrelated to CFI ratings (Freedman, Stumpf, & Aguanno, 1979). However, instructor presentation, ability, and concern for his/her students did correlate with CFI ratings.

References

Freedman, R. D., Stumpf, S. A., & Aguanno, J. C. Validity of the course-faculty instrument (CFI): intrinsic and extrinsic variables. Educational and Psychological Measurement, 1979, 39(1) 153-158.

Freedman, R. D. & Stumpf, S. A. Student evaluations of courses and faculty based on a perceived learning criterion: Scale construction, validation, and comparison of results. Applied Psychological Measurement, 1978, 2, 189-202.

Freedman, R. D., Stumpf, S. A., & Krieger, K. M. Business students' perceptions of subject matter, their learning, evaluation of courses, and school comparisons. Presented at the 38th Annual Meeting of the Academy of Management, San Francisco, California, August, 1978. (NYU/GBA WP#78-22).

Stumpf, S. A. & Freedman, R. D. Expected grade covariation with student ratings of instruction: Individual versus class effects. Journal of Eduational Psychology, 1979, 71, 293-302.

Stumpf, S. A., Freedman, R. D., & Aguanno, J. C. A path analysis of factors often found to be related to student ratings of teaching effectiveness. Research in Higher Education, 1979, (in press).

Stumpf, S. A., Freedman, R. D., & Krieger, K. M. Validity extensions of the course-faculty instrument (CFI). Research in Higher Education, 1979, (in press).

Title: Arizona Course/Instructor Evaluation Questionnaire (CIEQ)

Contact: Lawrence M. Aleamoni, Director
Office of Instructional Research and Development
The University of Arizona
1325 E. Speedway Boulevard
Tuscon, Arizona 85721
(602) 626-4488

Format

The CIEQ rating form is available on a computer scorable answer sheet which is divided into five sections. The first section elicits student background information including: student level, whether the student is taking the course on a pass/fail basis, whether the student is taking the course as an elective, student sex, expected grade, the proportion of students taking the course as part of their major, and the semester in which the evaluation takes place. The second section consists of three general items which elicit student responses to the course content, the instructor, and the course in general. Ratings in this section are made on a six-point scale ranging from excellent to very poor. Section three includes twenty-one statements which represent five subscales or factors labeled as: General Course Attitude, Method of Instruction, Course Content, Interest, and Attitude. A sixth scale, Instructor, provides scores for all items combined. Items are rated on a four-point scale ranging from agree strongly to disagree strongly. The fourth section provides space for 42 optional items if the instructor wishes to

include any additional items. These items may either be selected from an item catalogue which is part of the CIEQ system or written by the instructor. The final section allows for open-ended responses to questions on course content, the instructor, course objectives, papers and homework, examinations, suggested improvements, and an evaluation of the course based upon student satisfaction with the course and student perceptions of its value as an educational experience.

Results

The results of the CIEQ are presented on computer output in four parts. The first part presents course identification and student background information. The second part lists results for the three general items. Given are the proportion responding to each item alternative (and the proportion not responding), the mean, and the standard deviation for each item. The third part lists each of the twenty-one standard items and gives the proportion responding to each alternative, the most favorable response, the mean response, the standard deviation, and the college-wide norm decile (a comparison of the mean response with those obtained throughout the college or university) for each item. The final part lists the responses to the five subscales. Included are the percent responding, the mean response, the standard deviation, the reliability coefficient (based upon an internal consistency calculation), and a variety of normative comparisons. These comparisons include: the rank norm (a comparison of the course with all courses given by instructors at the same rank), the level norm (a comparison of the course with all courses at the same course level), the institution norm (a

comparison of the course with all courses at the university), the college norm (a comparison of the course with all other courses in the appropriate college within the university), the overall norm (a comparison of the course to all the courses throughout the U.S. which have used the CIEQ), and the department norm (a comparison of the course with all other courses in a particular department).

Special Features

The optional item catalogue (Aleamoni & Carynnk, 1977) contains 350 items divided into twenty categories. The Results Interpretation Manual (Aleamoni, undated) provides information on scale development and validation, recommended uses and administrative procedures, description and interpretation of results, and decile norm cutoff scores for seven data bases.

Development & Validation

The CIEQ was developed in 1975 through an analysis of the Illinois Course Evaluation Questionnaire (CEQ). The original CEQ was based on an initial pool of over 1000 items collected in the early 1960's, reduced and refined by a variety of techniques, including factor analysis, to a form containing 50 items (Aleamoni & Spencer, 1973). The current version (Form 76) utilizes normative data from roughly 1600 course sections at the University of Arizona and the University of Illinois at Urbana-Champaign and 6800 sections from other U.S. institutions gathered from 1972 through 1978. Internal consistency reliability coefficients for the five subscales range from .81 to .94 (Aleamoni, undated). Test-retest reliability ranges from .92

to .98 for the subscales and the total and from .81 to .94 for individual items (Gillmore, 1973). Aleamoni (1978) reviews several studies of the CEQ which he claims are generalizable to the CIEQ. He reports that the CIEQ is not affected by sex, term, curriculum, class size, instructor rank, required/elective, major/minor, student status, pass/fail, expected grade, and final grade. In addition, the ratings of colleagues and trained judges appear to correlate with CIEQ student ratings (Aleamoni, 1978).

References

Aleamoni, L. M. Development and factorial validation of the Arizona course/instructor evaluation questionnaire. Educational and Psychological Measurement, 1978, 38, 1063-1067.

Aleamoni, L. M. Arizona course/instructor evaluation quetionnaire: Results interpretation manual. Tucson, Arizona: University of Arizona, Office of Instructional Research and Development, undated.

Aleamoni, L. M. & Carynnyk, D. B. Optional item catalogue (revised). (Information Memorandum No.6) Tucson, Arizona: University of Arizona, Office of Instructional Research and Development, 1977.

Aleamoni, L. M. & Spencer, R. E. The Illinois course evaluation questionnaire: A description of its development and a report of some of its results. Educational and Psychological Measurement, 1973, 33, 669-684.

Gillmore, G. M. Estimates of reliability coefficients for items and subscales of the Illinois course evaluation questionnaire (RR 341). Urbana, Illinois: University of Illinois, Measurement and Research Division, Office of Instructional Resources, 1973.

Title: The Endeavor Instructional Rating System (Endeavor)

Contact: Peter W. Frey
Endeavor Information Systems Inc.
2407 Prospect Ave.
Evanston, Illinois 60201
(312) 869-1034

Format

The Endeavor rating form is a computer scorable, "porta-punch" answer card which consists of seven items designed for summative purposes. Each item deals with a separate course or teacher characteristic: course workload, course planning, class discussion, instructor helpfulness, presentation clarity, grading accuracy, and course contribution to increased knowledge. Responses to each item are on a seven point scale ranging from never to always. The items represent two global factors: Pedagogical Skill and Rapport. The Endeavor rating form can be administered in five to ten minutes but does not provide space for additional questions.

Results

The results for each instructor are provided in a five page computerized report. The first page contains a tabular comparison of the instructor's scores on each item with those for all other instructors in the same thematic group (e.g., history professors). Group means are also provided for each item. The next page provides a scatterplot of scores on the two global factors, Pedagogical Skill and Rapport. Each instructor is

designated by a letter on a graph defined by the two axes which represent the two global factors. The position of the letters in relation to the axes indicates the mean ratings which were received on each factor. Page three provides a detailed breakdown of the instructor's ratings which includes a frequency count, means, and standard deviations for each item. In addition, global ratings are calculated for the Pedagogical Skill and Rapport factors. The fourth page provides frequency histograms for each item. The final page tables the mean ratings for each department and the institutional means for each of the seven items.

Special Features

Endeavor is a commercial rating package whose forms are prepared and scored by Endeavor Information Systems. The services of consultants are also available. The cost of the system varies with the number of forms used by the institution from $.16 per form for 1000 or less to $.09 per form for more than 6000. Four copies of the results are provided for each course evaluated. Processing time is two weeks or less. A 16-page user's handbook is also available which includes: (a) a brief introduction to some issues in the development and use of student rating forms for promotion and tenure decisions; (b) a description of the development of the Endeavor system; (c) detailed instructions for the administration and collection of Endeavor rating forms; (d) a summary of research on the reliability and validity of the Endeavor system; and (e) guidelines for interpreting the ratings.

Development & Validation

The current seven item rating form is a refinement of a number of much longer rating forms dating back to 1971. These earlier forms were reduced because of student complaints of length and item redundancy. The current version was adopted in 1974. Reliability data indicates that Pedagogical Skill is quite stable over time (r=.90) while Rapport shows a correlation of .38. Test-retest reliabilities for the individual items range from .30 to .87, with items representing Skill far more stable than those representing Rapport. Pasen (1977) reported that the mean scores on the two global factors were virtually identical over a one year period. Frey (1975) reported interrater agreement of .61 for Pedagogical Skill and .32 for Rapport. In studies in which student learning was used as the criterion, the Pedagogical Skill measure showed a median correlation of .81 while the Rapport measure showed one of .29 (Frey, 1978).

Research on the Endeavor form has explored the relationship between ratings and a number of potential sources of bias. These include: average grade in class, individual student grade, quantitative SAT score, student class level, instructor rank, citation count, student major, class size, and administration instructions. Frey (1978) found that average grade in class did not correlate with Skill items but did correlate positively with Rapport items. Student class level was found to be related to most Endeavor items (Frey, Leonard & Beatty, 1975). Instructor rank was negatively correlated with Rapport ratings but positively correlated with Skill ratings (Frey, 1978). Class size (Frey, 1978), student GPA (Frey et.al., 1975), and student sex were unrelated to Endeavor ratings.

References

Frey, P.W. A two-dimensional analysis of student ratings of instruction. Research in Higher Education, 1978, 9, 69-91.

Frey, P. W. Validity of student instructional ratings: Does timing matter? Journal of Higher Education, 1976, 47, 327-336.

Frey, P. W., Leonard, D. W., & Beatty, W. W. Student ratings of instruction: validation research. American Educational Research Journal, 1975, 12, 435-447.

Pasen, R. M. The differential effect of grade, sex, and discipline on two global factors: A within-class analysis of student ratings of instruction. Ph.D. Dissertation, Northwestern University, June 1977.

Pasen, R. M., Frey, P. W., Menges, R. J., & Rath, G. J. Different administrative directions and student ratings on instruction: Cognitive versus affective states. Research in Higher Education, 1978, 9, 161-168.

Title: The University of Washington Instructional Assessment System (IAS).

Contact: Gerald M. Gillmore
Educational Assessment Centre
University of Washington
1400 Campus Parkway
PB-30
Seattle, Washington 98195
(206) 543-1170

Format

The IAS consists of six rating forms, each tailored to one of six basic course types: Form A is designed for small lecture-discussion type courses; Form B for large lecture type courses; Form C for seminar type courses; Form D for problem-solving type courses; Form E for skills oriented or practicum courses; and Form F for quiz sections which are offered by graduate teaching assistants in conjunction with a large lecture section. All forms are subdivided into five evaluative sections plus items assessing student background characteristics (e.g., student level, expected grade, program requirements). Except for responses to items in the final section, all student responses are marked on a computer scorable answer sheet. All items in the first three sections are rated on a six-point scale ranging from excellent to very poor. Section 1 is designed to provide general evaluative information. It consists of four items which ask for overall ratings of the course, content, instructor's contribution, and instructor's effectiveness. These items appear on all forms allowing comparisons to be made

college-wide as well as department-wide. Section 2 is designed to provide diagnostic feedback to the instructor. It consists of 11 items which ask for ratings on specific instructor characteristics. Items are common within course type, allowing normative comparisons with other, similar courses. Examples of characteristics typically included are: organization, voice clarity, clarity of explanations, presentation of alternative explanations, use of examples, quality of issues raised, knowledge, enthusiasm, encouragement given to students, ability to answer questions, and availability for help outside of class. Section 3 is designed to provide information about the course and the instructor to students considering enrolling in the course. Advanced approval must be given by the instructor before the Section 3 results are made public. It consists of seven items, common to all forms, dealing with use of class time, instructor's interest in his/her students, amount learned, relevance of course content, grading, workload, and clarity of course requirements. Section 4 provides space for eight instructor-designed questions. Response alternatives provide for ratings on either the six-point scale used in the previous sections or a ten-point scale. Section 5 is an optional open-ended comment sheet which asks students to describe the aspects of the course which they felt were especially good and to indicate what changes could be made to improve the course.

Results

A one page computer output summarizes the results of an instructor's evaluations, excluding student responses to the open-ended questions from Section 5 which are passed directly to

the instructor. Frequency distributions of student responses to the background questions are listed. However, these items are not used to adjust rating scores or create separate norm groups. For all evaluation items, frequency distributions and mean ratings are provided. For Section 1 items, each mean item score and the mean of the four items are placed in a centile rank for the form used and all forms combined. For each Section 2 and 3 item, the percentage of students who responded below "Good" is listed. The relative rank of each item mean within a section and for the particular instructor is given. The reverse side of the output provides a brief guide to interpreting the results.

Special Features

The IAS is designed to provide three sets of information; general information on course rating for adminstrators, specific-diagnostic information of instructor and course characteristics for instructors, and information to students who are considering taking the course. The different forms allow for the use of more specific or relevant items to suit different course types. IAS forms cost 2.5 cents each, regardless of the number purchased. Processing is available separately at a cost of 5 cents per form plus 75 cents per class. This includes two copies of every class summary output. Additional copies cost 10 cents each.

Development & Validation

The IAS was developed at the University of Washington by Gerald Gillmore and his associates in 1974. In its original form, it consisted of five forms (each dealing with a different

course type) and three basic sections (for administrator, instructor, and student information). The sixth form was added later for use with quiz sections and the forms were expanded to allow room for instructor-designed questions. Statistical analyses of the IAS are provided in Gillmore (1975) which is based on the system's first year of use. The analyses of all items from each form include: means, standard deviations, interitem correlations, and interrater reliabilities for different class sizes (N=1,10,& 40). Interitem correlations for items common to all forms range from .54 to .95. Correlations between nonevaluative variables and ratings were also computed. Class size, expected grades, desire to take the course prior to enrolling, and the evaluation form used were found to relate to ratings to a nontrivial extent. Reliabilities vary according to class size, from .15 to .34 for single raters to .88 and above for classes of 40 students.

References

Gillmore, G. M. A brief description of the University of Washington Instructional Assessment System. (EAC Report 276). Seattle: University of Washington, Educational Assessment Center, 1974.

Gillmore, G. M. Statistical analyses of the data from the first year of use of the student rating forms of the University of Washington Instructional Assessment System (EAC Report 503). Seattle: University of Washington, Educational Assessment Center, 1976.

Title: Instructor and Course Evaluation (ICE)

Contact: Thomas D. Purcell, Director
Instruction Evaluation Office
Southern Illinios University
at Carbondale
Carbondale, Illinois 62901
(618) 536-5593

Format

The rating form consists of five parts and a student biographical section used to develop appropriate norm groups. The form is designed to be scored by an optical scanning device for computer processing. Part I contains 19 items dealing with specific instructor characteristics such as preparation, grading, clarity, organization, enthusiasm, and availability for help. Also included is a general item assessing overall teaching effectiveness. Each item is scored on a five-point scale ranging from exceptional to improvement definitely needed. Part II contains 20 items dealing with specific course characteristics such as organization, content, interesting material, texts and examinations. This section also includes an overall rating of the course. Items are scored on a five-point scale ranging from strongly agree to strongly disagree. Part III asks each student to rate the progress he/she has made toward course objectives in this course compared to other courses. It consists of 10 items dealing with specific course objectives including gaining of factual knowledge, learning of fundamental theories, developing creative capacities, and oral expression. Items are scored on a five point scale ranging from exceptional progress to no

progress. The items in Part IV deal with the reasons behind the student's taking the course. These include interest level, expected grade and motives in taking the course as well as one item on the adequacy of the rating form for the course. These items are rated on a Yes-No basis. Part V allows space for optional instructor-designed items which have up to five response alternatives.

Results

The results are presented to the instructor in a computer output divided into six parts plus an introductory section summarizing student responses to the background questions. Parts I - V of the output list the percentage of student responses to each item alternative and item means. Except for Parts IV and V (student self-ratings and optional items, respectively) decile rankings of each item mean are given in relation to each instructor's department, college, and university. Total score means for the four ICE factors are reported with the university-wide decile rankings adjusted for course level and the proportion of students taking the course as a requirement. Part I summarizes the results for the ICE factors related to Instructor Characteristics, Personal-Interpersonal Skills, and Course Structure by Instructor. Part II summarizes the course factors of Quality and Difficulty. Part III lists Progress on Objectives. Part IV indicates the five items on which the instructor scored best and the five items on which he/she scored poorest for both instructor and course characteristics.

Special Features

Accompanying the ICE is an interpretive manual (Miller, 1978) which serves several functions. First, it explains the computer output summarizing the instructor's evaluations. Second, it briefly describes instructor, student, course, and curriculum characteristics related to instructional outcomes such as student achievement and attitudes. Third, it discusses instructor and course characteristics affecting ICE outcomes. Specific suggestions for teaching improvement are provided by describing teaching according to each of the items comprising the four factors of the ICE. A fourth section discusses aids to teaching and course improvement such as videotaped feedback, behavior modeling, speech training, and test construction. The appendices provide ICE factor loadings and summary statistics broken down by student biographical items.

Development & Validation

The ICE represents a revised version of the Instructional Improvement Questionnaire (IIQ). The revisions consist of items on course objectives (Part III) and student self-ratings (Part IV). Course objective items are intended to measure both cognitive and affective goal achievement. The 40 ICE items on instructor characteristics (Part II) are identical to IIQ items. The IIQ items were intended to have content validity for students, internal consistency, reliability of factor scores, test-retest reliability, factor structure stability across instructors and within instructors across time, and the ability to discriminate among instructors. In addition, a positive relationship between student ratings and achievement was seen as

desirable. Research on the IIQ is summarized in a series of published and unpublished reports.

Development of the IIQ (summarized by Pohlmann & Elmore, 1975) began with a review of existing rating forms and the generation of an initial item pool. The pool was reviewed by a committee consisting of students, faculty, and measurement specialists. Students and faculty reactions to the trial forms lead to the final development of the 20 course and 20 instructor items.

Miller (1978) summarizes a factor analysis of class means of the 40 IIQ items found on the ICE. The four factors accounted for 63.06% of total variance. Item factor loadings are moderate with correlations ranging from .25 to .79.

Pohlmann (1973) investigated the reliability of IIQ subscales. Internal consistency coefficients (Cronbach's alpha) ranged from .62 to .93. Three month test-retest correlations ranged from .67 to .76.

The relationship between course characteristics and IIQ results was studied by Pohlmann (1975). Two class characteristics, expected grade and percent of students taking the course as an elective, were positively and moderately related to IIQ scores. Elmore and LaPointe (1974) found no significant relationship between IIQ ratings and faculty and student sex. However, teachers viewed by students as warm and interested in students received high IIQ ratings (Elmore & LaPointe, 1974).

References

Elmore, P. B. & LaPointe, K. A. Effects of teacher sex and student sex on the evaluation of college instructors. Journal of Educational Psychology, 1974, 66, 386-389.

Elmore, P. B. & LaPointe, K. A. Effect of teacher sex, student sex, and teacher warmth on the evaluation of college instructors. Journal of Educational Psychology, 1975, 67, 368-374.

Miller, W. G. Guidelines for interpreting results of the instructor and course evaluation. Carbondale, Illinois: Southern Illinois University, Instructional Evaluation Office, 1978.

Pohlmann, J. T. Evaluating instructional effectiveness with the instructional improvement questionnaire. Technical Report 72-73. Carbondale, Illinois: Southern Illinois University, Instructional Evaluation Office, July, 1973.

Pohlmann, J. T. A multivariate analysis of selected class characteristics and student ratings of instruction. Multivariate Behavioral Research, 1975, 10, 81-91.

Title: Instructor and Course Evaluation System (ICES)

Contact: Dale C. Brandenburg
Office of Instructional Resources
Measurement and Research Division
307 Engineering Hall
Urbana, Illinois 61801
(217) 333-3490

Format

The rating form consists of three general core items dealing with course content, the instructor, and the course in general. These items are scored on a six point scale ranging from excellent to very poor. The form also provides space for up to 23 items which are chosen from a computer-stored item bank of over 600 items. The item bank catalogue lists the items according to general concept and specificity in six groups: Course Management, Student Outcomes of Instruction, Instructor Characteristics and Style, Instructional Environment, Student Preferences for Instruction/Learning Style, and Specific Instructional Setting. Tailor-made rating forms can be developed by departments (Option 1) or by individual instructors (Option 2). Alternately, sets of complete forms for specific course types and specific course areas are available (Option 3). All non-core items are scored on a five-point scale with only the endpoints labeled. These labels vary according to the item stem. Space is provided on the reverse side of the rating form for open-ended responses to four questions dealing with instructor strengths and weaknesses, the most beneficial aspect of the course, how the course might be improved, grading procedure and

two optional instructor-designed questions. The form also requests student information on their opinion of the instructor prior to taking the course, their sex, whether the course is required or an elective, whether the course is in their minor or major, their class level, and their expected grade.

Results

The computerized output varies somewhat depending on the rating form options selected. However, there are two sections of results reported regardless of the options selected. The first section summarizes student demographic data for students who completed ICES questionnaires. The second section summarizes the global item results. These include means, standard deviations, and frequency distributions for the three core items. The instructor's mean score on each core item is compared to two groups. Group 1 is one of six university norm groups, where the groups were determined on the basis of course status (mostly required, mixed, mostly elective) and instructor status (teaching assistant, faculty member). Group 2 is the instructor's department. Comparisons are not made only in terms of relative placement of the instructor's mean rating: a range of scores is presented which takes into account the unreliability of the items. Department core item (Option 1) information parallels that presented for the core items except that only departmental comparisons are given. Option 2 items (instructor selected) results contain no normative information. Complete form (Option 3) results are reported with items first and subscales second. Summary statistics are provided along with normative comparisons for all other instructors using the form.

Special Features

The ICES item bank catalogue is the first in a series of ICES newsletters which describe the rationale and uses of ICES. Because of the nature of the ICES system, detailed descriptions of teaching dimensions or specific suggestions for teaching improvement are not provided. Tailor-made and computer scorable rating forms can be easily generated by completing an evaluation request form. External users must order ICES questionnaires from the University of Illinois and return them for processing. Two copies of a faculty report will be returned with local institutional norms provided for the three global items and any items selected as part of an institutional core. The cost is $.20 or less per questionnaire depending on the number ordered.

Development & Validation

The ICES item bank is a set of items culled from other rating forms along with items generated by the authors. Validation of the item classification scheme for ICES is described in Brandenburg, Derry, and Hengstler (1978). Subsets of these items were initially classified in two ways: content categories (according to previous factor analytic studies) and item specificity (as global, general concept or specific). Factor analyses showed that the items could be grouped into two large domains: Influence (which consists of items related to student-perceived outcomes, instructor communication skills, and instructor/course stimulation) and Security (which consists of items related to course management/structure and instructor warmth/concern). Brandenburg (1979) describes some other statistical properties of ICES items. Item reliabilities for

global and general concept items typically range between .69 and .91 with a median of .80 for each subgroup, while reliabilities for specific items fluctuated more and were lower (.70). Within-class variances among student responses to all items were between .67 and .87. Brandenburg, Slinde, and Batista (1977) provide a rationale for the required/elective and rank breakdown utilized in ICES.

References

Brandenburg, D. C. Some statistical properties of item specificity. Paper presented at the annual meeting of National Council for Measurement in Education, San Francisco, California, April 1979.

Brandenburg, D. C., Derry, S., & Hengstler, D. D. Validation of an item classification scheme for a student rating item catalog. Paper presented at annual meeting of National Council for Measurement in Education, Toronto, Canada, March 1978.

Brandenburg, D. C., Slinde, J. A., & Batista, E. E. Student ratings of instruction: Validity and normative interpretations. Research in Higher Education, 1977, 1, 67-78.

Braskamp, L. A., Ory, J. C., & Pieper, D. M. Student written comments: Dimensions of instructional quality. Research Report #368. Urbana, Illinois: University of Illinois, Office of Instructional Resources, July 1979.

Ory, J. C., Brandenburg, D. C., & Pieper, D. C. Selection of course evaluation items by high and low rated faculty of varying academic rank. Paper presented at the annual meeting of National Council for Measurement in Education, San Francisco, April 1979.

Ory, J. C., Braskamp, L. A., & Pieper, D. M. The congruence of student evaluative information collected by three methods. Research Report #369. Urbana, Illinois: University of Illinois, Office of Instructional Resources, July 1979.

Title: Instructional Development and Effectiveness Assessment System (IDEA)

Contact: William E. Cashin
Center for Faculty Evaluation
and Development
Kansas State University
1627 Anderson Avenue
Box 3000
Manhattan, Kansas 66502
(800) 255-2757 or (913) 532-5870

Format

The IDEA system requires the instructor to describe his course objectives prior to the administration of the rating form. The instructor is asked to rate the importance, on a three point scale (essential, important, of minor importance), of each of ten IDEA objectives which fall into three broad areas: Subject Matter Mastery, General Skills Development, and Students' Personal Development. The importance the instructor assigns to each objective is taken into account in the tabulation of results. The rating form is divided into seven parts, with student responses recorded on specially designed optical mark reader cards. The first section consists of 20 items which deal with instructor characteristics. Students are asked to describe the frequency of each of 20 teaching procedures such as promotion of teacher-student discussion, instructor enthusiasm, expressiveness and variety of voice tone, and clarity of course objectives. Items are scored on a five-point scale ranging from hardly ever to almost always. The second section deals with the students' evaluation of their progress on 10 course objectives including

gaining factual knowledge, learning fundamental principles, developing creative capacities and discovering implications of the course material for understanding oneself. Students are asked to rate the progress made in the course on each objective as compared with the progress made in other courses. Each item is scored on a five-point scale ranging from low (lowest 10 per cent of courses taken) to high(highest 10 per cent of courses taken). The third section deals with four course characteristics: amount of reading, amount of work in other assignments, difficulty of subject matter, and degree to which the various elements of the course seemed to be related. Characteristics are compared to other courses on a five-point scale ranging from much less than most courses to much more than most. The next section includes a self-rating of student's attitudes and behaviors in the course. Items assess the amount of work put into the course, desire to take the course, desire to take another course with the same instructor, feeling toward the field of study as a result of having taken the course, and consideration given to the questions on the course evaluation form. Each item is scored on a five-point scale from definitely false to definitely true. The fifth section deals with demographic data such as student sex, age, and years at school as well as questions on expected grade, number of courses evaluated using the IDEA form, and how well the form permitted the student to accurately describe the instructor and the course. The sixth section is for instructor-designed multiple choice questions. Finally, students may use the back of the response card for open-ended comments.

Results

The IDEA report consists of seven parts plus course identification data (e.g., number of students enrolled, class size, student motivation). The first three parts correspond to the second, third, and fourth sections of the rating form. The tabulation of results for each item includes the percentages responding at each level, mean rating, and a comparison of the mean with all courses rated and with courses of similar class size (four categories) and student motivation (five levels). These comparative data are based on the evaluations of approximately 23,000 courses from over 150 colleges and universities using the IDEA system. Part I (Evaluation) addresses the effectiveness of the instructor's teaching by presenting comparative rankings on a five point scale (low to high) only for those objectives which the instructor described as relatively important for that course. In addition, an overall evaluation is presented which represents progress on relative objectives in comparison to national norms. Part II (Course Description) provides information on how students described selective aspects of the course. Part III (Student Self-Ratings) describes certain characteristics of the students in the course. Part IV (Methods) summarizes responses to items dealing with instructor's teaching procedures. The 20 items are grouped under the four headings of Involving Students, Communicating Content and Purpose, Creating Enthusiasm, and Preparing Examinations. Percentages responding at each level and mean rating are presented for each item along with a difference score which indicates how the mean score differs from classes of similar size and student motivation level. This difference score is

translated into a rating of low, medium or high to allow for easier interpretation. Part V (Additional Questions) provides a tabulation of responses to instructor-designed questions. Part VI (Diagnostic Summary) attempts to identify the modifications necessary to an instructor's teaching methods in order to improve teaching effectiveness. The summary lists the course objectives described as important or essential by the instructor. If student ratings of progress on each of these objectives is average or below, the teaching methods most needing attention are listed. The item numbers of the teaching methods which are related to progress on each objective are listed if the instructor received a low rating on those methods. Part VII (Summary Profile) provides a graphic representation of three teaching outcome measures and four measures of teaching methods in comparison to similar courses and all courses.

Special Features

The IDEA system is a commercial rating package. Charges range from $.33 per form for forms plus scoring (100 to 3499 forms) to $.20 per form for forms plus scoring (100,000 or more forms). If over 10,000 forms are ordered, the Center for Faculty Evaluation and Development will provide some free on-campus consultation. Forms must be ordered from the Center and returned to them for processing. Institutions will receive three copies of IDEA computer reports plus one interpretive guide for each class evaluated. When fifteen or more classes are evaluated at once, an institutional summary is also provided. In addition to the interpretation manual, the Center publishes Exchange, an occasional newsletter, and has available a series of technical

and nontechnical publications on topics in faculty evaluation and development. Workshops and other services can also be arranged.

Development & Validation

The development and initial validation of the IDEA system is described by Hoyt and Cashin (1977). Items on instructor objectives were originally formed from taxonomic classifications and the factor analytic work of Deshpande and Webb. These items were subsequently critiqued and revised by award winning teachers, faculty-student committees, and early users of IDEA. The 20 teaching method items now in use are based on an initial pool of items culled from existent questionnaires. This initial pool was reduced by examining: item redundancy, item relationships with an overall effectiveness measure, responses of outstanding teachers and faculty-student committees, and item relationships with progress on objectives. Items on course management and student characteristics were also included because it was felt that program objectives might be contingent on such factors. The reliability of individual method scales ranged from .81 to .94, averaging .87 for classes of 25 students. The reliability of progress items was similar, ranging from .84 to .90, also averaging .87. The validity of student ratings of progress, the relationships between teaching methods and objectives and the influence of class size and student motivation were also explored. Student reports of progress on objectives were found to be positively related to instructor ratings of importance of objectives. Multiple regression analyses showed that each of the six teaching method scales contributed to the prediction of at least one progress-on-objectives ratings. Item

analyses also revealed that specific student ratings of instructor behaviors were related to progress but that this varied with class size. Student motivation was also found to be strongly related to progress ratings. More recent technical reports (e.g., Cashin & Perrin, 1978) provide additional data on reliability and validity as well as describe the computational procedures and comparative data bases used in producing reports for the latest version of IDEA.

References

Cashin, W. E. & Perrin, B. M. Description of the IDEA system data base--1978-79 (IDEA Technical Report No. 4). Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development, December 1978.

Cashin, W. E. & Slawson, H. M. Description of the IDEA system data base--1976-77 (IDEA Technical Report No. 2). Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development, April 1978.

Cashin, W. E. & Slawson, H. M. Description of the IDEA system data base--1977-78 (IDEA Technical Report No. 3). Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development, December 1977.

Hoyt, D. P. & Cashin, W. E. Development of the IDEA system (IDEA Technical Report No. 1). Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development, March 1977.

Title: Instructor Designed Questionnaire (IDQ)

Contact: James A. Kulik
Center for Research on Learning
and Teaching
University of Michigan
109 E. Madison
Ann Arbor, Michigan 48109
(313) 764-0505

Format

The IDQ rating form consists of five core items and up to twenty-five other items plus space for two essay questions. A computer-scorable answer sheet is available for student responses to multiple choice items which must all be made on a five-point scale ranging from strongly agree to strongly disagree. The core items deal with overall ratings of the course and the instructor, instructor's ability to motivate students, students' ratings of their performance in the course, and students' desire to take the course. The non-core items must be selected from an item catalogue except for three items which may be originals. The catalogue describes a computerized bank which includes roughly 200 multiple choice items organized under three major headings: student development, instructor ratings, and course elements. In addition, the catalogue lists 11 possible essay questions and a set of eight student counseling items which the instructor may select if he wishes some of his ratings to be made available for student use. A subset of required items for all instructors within a department may also be set.

Results

The single-page computer output returned to the instructor summarizes the results of the multiple choice responses. Frequency distributions and medians are reported for each item. For every item except originals, normative data are provided based on other users of the item. Median scores are reported for three groups: lowest 25%, middle 50%, and upper 25%.

Special Features

If the instructor has chosen to include the student counseling items, a second report is prepared. This report summarizes the results of the university-wide and student counseling items only. It may be used to advise students on course selection.

Development & Validation

The IDQ is modeled after Purdue's Cafeteria system which also includes a core set of items and allows the instructor to select additional items from a catalogue. The IDQ includes items covering various cognitive objectives of teaching as described in Bloom's taxonomy: knowledge, concepts, application, analysis, synthesis, and evaluation. Items reflecting affective objectives of instruction, as outlined by Krathwohl, are also included. Items are listed which describe the six major areas of teaching behavior: Teaching Skill, Rapport, Interaction, Feedback, Organization, and Difficulty. Other items were included to focus on specific course elements. Specific validation research on the IDQ is not yet available although prior validation research conducted by CRLT staff (McKeachie and others) is relevant to particular IDQ items.

References

Kulik, J. A. Student reactions to instruction. Memo to the faculty, (No. 58), October, 1976. Ann Arbor, Michigan: University of Michigan, Center for Research on Learning and Teaching.

Kulik, J. A. Using CRLT's instructor-designed questionnaire to improve instruction. Ann Arbor, Michigan: University of Michigan, Center for Reaseach on Learning and Teaching, 1978.

Title: The Student Instructional Report (SIR)

Contact: Nancy Beck or Eldon Parks
Student Instructional Report
Educational Testing Service
Box 2813
Princeton, New Jersey 08540
(609) 921-9000

Format

The SIR rating form consists of 39 core items which are divided into three sections with space for up to ten instructor-prepared supplementary questions. The form is printed on both sides of a computer answer sheet in various formats (IBM 1230, IBM 3881, Digitek). Section I, which focuses on instructor characteristics, is made up of 20 items dealing with such factors as use of class time, stimulation of interest, and concern for students. Items are scored on a four-point scale ranging from strongly disagree to strongly agree. Students have the option of choosing a fifth response choice, NA, if the question is not applicable or if they feel they are unable to give a knowledgeable response. Section II consists of 10 items, some of which deal with course characteristics such as level of difficulty, workload and pace at which the material is presented. Other items deal with student characteristics such as expected grade, sex, and reasons for taking the course. Section III contains items on course variables such as texts, exams, and laboratories as well as an overall course rating item and an item which assesses the instructor's overall effectiveness as compared to other instructors. All but this last item are scored on a

five-point scale ranging from poor to excellent with the students also having the option of responding NA if they choose. The optional instructor-designed questions in Section IV can use up to 10 response alternatives per question.

Results

The results are summarized in a three-page report which provides a frequency distribution of students responses for each item and the item mean, where appropriate. When class size is small and/or the percentage of students responding is low, responses may be flagged or not tabulated. This reduces the probability of interpreting unreliable data. For 29 items percentile equivalents are also provided. These ratings are based on a comparison of the instructor's average score on each item with national norms gathered from SIR administrations at either two-year colleges and technical institutions or four-year colleges and universities. Additional comparative data guides have been prepared separately for both institution types. Each guide contains data analyzed for: specific institution types, class size, class level, class type, and subject area.

Special Features

SIR is a commercial rating system; the questionnaire must be purchased from Educational Testing Service. However, institutions may score their own forms if they wish. Costs vary according to the number of rating forms ordered. For example, the cost of up to 5000 answer sheets is 12 cents per form plus 28 cents each for scoring, while orders in excess of 60,000 run 6 cents per form plus 18 cents for proccessing. This cost is

cumulative, thus the cost of using SIR will decrease from year to year. Three copies of each class report are returned within three weeks of receipt of the answer sheets by ETS. Summary reports are available in combinations requested by the institution (e.g., departmental or institution). The SIR package also includes an instructor self-evaluation form. Special student rating forms are available in Spanish, and in English and French for Canadian institutions. The SIR Bulletin is a semiannual newsletter for institutional users.

Development & Validation

The preliminary SIR form contained 112 items divided into five groups: Course Organization & Content, Instructor-Student Relations, Communication, Assignments & Evaluation, and Student's Involvement in the Course (Centra, 1972). This form was revised twice on the basis of factor analytic studies and faculty opinions of the items until 39 items remained. Centra (1973) found item reliabilities to be generally above .70 when comparing within class variances to between class variances for 20 or more students per class. The test-retest reliability of the majority of the items was also found to be above .70. Principal axis factor analysis employing oblique promax rotation identified six factors among 31 relevant SIR items (Centra, 1973). These factors were labeled: Teacher-Student Relationship, Course Objectives and Organization, Reading Assignments, Course Difficulty and Workload, Lectures, and Examinations. Intercorrelations among the six factors suggested that the factors are reasonably interrelated. Centra (1973) also found that the rank order correlations of student and alumni SIR

ratings of the same teachers were significantly positive. Mean SIR ratings were correlated with exam scores in 72 sections of seven courses (Centra, 1976). In two of the courses, students were randomly assigned. Global ratings of teaching effectiveness and course value were the most highly related to exam scores. Ratings of Course Workload and Student/Teacher Relationship were not highly related to exam scores. Centra (1976) investigated possible bias in student ratings by considering a number of student, teacher, and course characteristics. The findings revealed that many of the variables have relatively weak relationships to ratings.

References

Centra, J. A. The Student Instructional Report: Its development and uses (SIR Report No. 1). Princeton, New Jersey: Educational Testing Service, 1972.

Centra, J. A. Two studies on the utility of student ratings for improving teaching. I. The effectiveness of student feedback in modifying college instruction. II. Self-ratings of college teachers: A comparison with student ratings (SIR Report No. 1). Princeton, New Jersey: Educational Testing Service, 1972.

Centra, J. A. Comparisons with alumni ratings, reliability of items, and factor structure (SIR Report No. 3). Princeton, New Jersey: Educational Testing Service, 1973.

Centra, J. A. Two studies on the validity of the Student Instructional Report: I. Student ratings of instruction and their relationship to student learning. II. The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness (SIR Report No. 4). Princeton, New Jersey: Educational Testing Service, 1976.

Title: Student Instructional Rating System (SIRS)

Contact: LeRoy A. Olson
Learning and Evaluation Services
Michigan State University
17 Morris Hall
East Lansing, Michigan 48824
(517) 355-1855

Format

The SIRS rating form is a machine scorable answer sheet which consists of twenty-eight standard multiple choice items, space for an additional seven multiple choice items, and an open-ended comment area on the back of the form. Except for two items, student responses must be made on a five-point scale ranging from strongly agree to strongly disagree. The multiple choice questions are divided into three sections under the following headings: Instruction, Student Background, and Laboratory/Recitation. The Instruction section is made up of twenty-one items dealing with instructor and course characteristics including concern for students, enthusiasm, and course organization. Space is provided for three additional items which the instructor may select from the SIRS Supplementary & Diagnostic Item Catalog. The Student Background section consists of four items dealing with students' reason for choosing the course, GPA, and number of other related courses taken. Space is also provided for two instructor-selected questions. The final section is only to be filled out if the course included a laboratory or recitation section. The three standard items deal with the overall quality of the laboratory or recitation,

while space is provided for two additional items.

Results

The SIRS report is a four page computer output divided into six sections: Identification, Information, Column Headings, Laboratory or Recitation, Instruction, Student Background, and SIRS Composite Profile. The report lists the percentage of student responses to each item alternative, item means, and standard deviations. Also provided for each item are percentile ranks for instructors teaching the same course, for instrucors teaching courses in the same department, and for all instructors in the college or university. A composite profile is included which lists means, standard deviations, and percentile ranks on the following five global factors: Instructor Involvement, Student Interest, Student-Instructor Interaction, Course Demands, and Course Organization.

Special Features

SIRS is not a commercial rating package, although the materials are available to other institutions and arrangements for processing forms can be made on an individual basis with the Learning and Evaluation Services. The SIRS Manual briefly describes the rating form and administration and scoring practices. It contains a guide to interpreting the computer output and student ratings generally. Appendices include more than 200 supplementary and diagnostic items for use as SIRS optional items and 20 suggestions for open-ended questions.

Development & Validation

The development of the 21 standard instructor rating items of the SIRS form is described in the Technical Bulletin (Davis, 1969). Initially, 250 items were constructed based on interviews with students and instructors in a wide range of courses. A sample of faculty and students were then asked to rate the items by responding to four questions about each item, such as whether they would include the item on a rating form. This led to the selection of 52 items for the preliminary form. Factor analyses using a varimax solution revealed a five factor solution from which the highest loading items were selected per factor (see also Research Report # 2, 1971). These 20 items plus a general affect item made up the final form. SIRS Research Report # 4 (1972) describes the relationship between student SIRS responses and both student characteristics and student achievement. Multivariate multiple regression analyses showed that three student characteristics (previous GPA, sex, and expected grade) were significantly related to SIRS composite profile scores. However, the correlations between profile scores and student characteristics were low, and the proportion of variance accounted for was small. SIRS composite profile scores were also found to be significantly related to student achievement variables. However, the proportion of variance accounted for was small. SIRS item reliabilities are reported by Showers (1974). Individualized item corrected split-half reliabilities for class sizes of 20 students averaged .75, ranging from .53 to .90.

References

Analysis of Responses for Winter Term, 1970 (SIR Research Report # 1). East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1971.

Davis, R. H. Student Instructional Rating System (SIRS) Technical Bulletin, East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1969.

Showers, B. SIRS Research Report #5: The effects of three kinds of response options on student ratings of instruction. East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1974.

Stability of Factor Structure (SIR Research Report #2). East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1971.

Student Characteristics (SIR Research Report # 4). East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1972.

Student Instructional Rating System (SIRS). (Testing and Evaluation Bulletin No. 9). East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1970.

Student Instructional Rating System Manual. East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1970.

Using the Student Instructional Rating System Report in the Decision-Making Process. East Lansing, Michigan: Michigan State University, The Office of Evaluation Services, 1972.

Title: Student Perceptions of Teaching (SPOT)

Contact: Rina R. Weerts
Evaluation and Examination Service
University of Iowa
300 Jefferson Building
Iowa City, Iowa 52242
(319) 353-5301 or (319) 353-4645

Format

The SPOT system allows the instructor to select up to 20 items from the SPOT item catalogue. These items are printed on one side of a computer scorable answer sheet. All items must be marked on a six point scale ranging from strongly agree to strongly disagree. The reverse side of the answer sheet may be used for up to eight instructor-designed questions. Student responses to these items may be on any scale with six alternatives or less. Space is also provided for student written comments.

Results

SPOT results are reported on a computer summary which lists for each item: the number of students responding, the percentage of students choosing each response alternative, and the item mean. In addition, for items selected from the SPOT catalogue for which sufficient normative data exist, item means are translated into one of three percentile ranks. The three groups used for comparisons are "Base", representing data from a 1974 SPOT administration, "Dept", representing data from all classes in the instructor's department, and "User", representing data

from all users at the university. Below each percentile rank a confidence interval is printed.

Special Features

The SPOT item catalogue lists in excess of 100 items organized into seven sections. The first four sections (Course Content, Structure, and Organization; Instructor's Behavior; Instructional Methods and Materials; and Outcomes of Instruction) include items particularily useful for traditional style courses. The supplementary items in the remaining sections (Laboratory Courses and Sections; Clinical Courses; and Production Courses) are for more specialized teaching situations. There is a uniform price for SPOT forms ($.05 each) and computer scoring ($.10 per form) regardless of the number ordered.

Development & Validation

An initial pool of more than 800 items was gathered from other rating forms and subdivided into four categories. The pool was reduced to 139 items by excluding redundant items, unclear items, overly specific items, etc. These items were submitted to the faculty and a sample of students at the University of Iowa in March, 1974. They were asked to rate each item for importance to learning and student ability to evaluate the characteristic studied. The results of this survey yielded an 80-item SPOT pool. Other items have been added to meet needs for courses and instructional situations.

Weerts and Whitney (1975) investigated the relationship between average student responses to the 80 items and selected student and instructor characteristics. In general, they found

small and inconsistent relationship between SPOT item scores and class size, class rank, required vs. elective course, percent of class responding, instructor rank, instructor sex, instructor age, and instructor graduate preparation. Average expected grade was substantially related (.30 or higher) to 31 of 80 SPOT items.

References

Student perceptions of teaching (SPOT): II. Asking the right questions. (Research Report No. 76). Iowa City, Iowa: University of Iowa, Evaluation and Examination Services, August 1974.

Student perceptions of teaching (SPOT): III. The 1978-1979 item pool. (Memo No. 26). Iowa City, Iowa: University of Iowa, Evaluation and Examination Services, August 1978 (revised).

Student perceptions of teaching (SPOT): IV. Baseline data for a representative sample of classes. (Summary Report No. 44). Iowa City, Iowa: University of Iowa, Evaluation and Examination Services, December 1976 (revised).

Weerts, R. R. & Whitney, D. R. Student perceptions of teaching (SPOT): V. Relationships between averaged student responses and selected course characteristics. (Research Report No. 78). Iowa City, Iowa: University of Iowa, Evaluation and Examination Services, June 1975 (revised).